

Bayesian Robust Quantile Regression

Mauro Bernardi^a, Marco Bottone^b, Lea Petrella^c

^a*Department of Statistical Sciences, University of Padova, Padua, Italy.*

^b*Bank of Italy and Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome, Rome, Italy.*

^c*Department of Methods and Models for Economics, Territory and Finance Sapienza University of Rome, Rome, Italy.*

Abstract

Traditional Bayesian quantile regression relies on the Asymmetric Laplace distribution (ALD) mainly because of its satisfactory empirical and theoretical performances. However, the ALD displays medium tails and it is not suitable for data characterized by strong deviations from the Gaussian hypothesis. In this paper, we propose an extension of the ALD Bayesian quantile regression framework to account for fat-tails using the Skew Exponential Power (SEP) distribution. Beside having the τ -level quantile as parameter, the SEP distribution has an additional key parameter governing the decay of the tails, making it attractive for robust modeling of conditional quantiles at different confidence levels. Linear and Generalized Additive Models (GAM) with penalized spline are considered to show the flexibility of the SEP in the Bayesian quantile regression context. Lasso priors are considered in both cases to account for shrinking parameters problem when the parameters space becomes wide. To implement the Bayesian inference we propose a new adaptive Metropolis–Hastings algorithm in the linear model and an adaptive Metropolis within Gibbs one in the GAM framework. Empirical evidence of the statistical properties of the proposed SEP Bayesian quantile regression method is provided through several example based on simulated and real dataset.

Keywords: Bayesian quantile regression, robust methods, Skew Exponential Power, GAM.

1. Introduction

Quantile regression has become a very popular approach to provide a wide description of the distribution of a response variable conditionally on a set of regressors. While linear regression analysis aims to estimate the conditional

URL: mauro.bernardi@unipd.it (Mauro Bernardi), marco.bottone@uniroma1.it (Marco Bottone), lea.petrella@uniroma1.it (Lea Petrella)

mean of a variable of interest, in quantile regression we may estimate any conditional quantile of order τ with $\tau \in (0, 1)$. Since the seminal works of Koenker and Bassett (1978) and Koenker and Machado (1999), several papers have been proposed in literature considering the quantile regression analysis both from a frequentist and a Bayesian points of view. For the former, following Koenker (2005) and the references therein, the estimation strategy relies on the minimization of a given loss function. From the Bayesian point of view Yu and Moyeed (2001) introduced the ALD as likelihood tool to perform the inference. After that a wide Bayesian literature has been growing on quantile regression and ALD see for example Dunson and Taylor (2005), Kottas and Gelfand (2001), Kottas and Krnjajic (2009), Thomson et al. (2010), Salazar et al (2012) Lum and Gelfand (2012), Sriram et al (2013) and Bernardi et al. (2015). Although the ALD is widely used in the Bayesian framework it has the main disadvantage of displaying medium tails which may give misleading informations for extreme quantiles in particular when the data are characterized by the presence of outliers and heavy tails. In fact the absence for the ALD of a parameter governing the tails fatness may influence the final inference. Recently Wichitaksorn et al. (2014) tried to generalize the classical Bayesian quantile regression by using some skew distributions obtained through mixture of scaled Normal ones. This class of distributions allows for different degrees of asymmetry of the response variable but they all impose a given structure of the tails. To overcome this drawback we propose an extension of the Bayesian quantile regression by using the Skew Exponential Power (SEP) distribution proposed in Zhu and Zinde-Walsh, 2009. The SEP distribution, like the ALD, has the property of having the τ -level quantile as the natural location parameter but it also has an additional parameter (the shape parameter) governing the decay of the tails. Using the proposed distribution in quantile regression we are able to robustify the inference in particular when outliers or extreme values are observed. In linear regression analysis several works have extensively considered the non skewed version of the SEP i.e. the Exponential Power distribution (EP), for the related robustness properties given by the shape parameter. Box and Tiao (1973) first show how to robustify the classical Gaussian linear regression model introducing the EP as distribution assumption for the error term. Choy and Smith (1997), explore the robustness properties of posterior moment based on the EP distribution, while Choy and Walker (2003) present further extension of the work of Choy and Smith (1997) introducing the case in which the shape parameter assumes values greater than two.

Finally, Naranjo et al. (2015) and Kobayashi (2016) consider the use of the SEP distribution in the regression and stochastic volatility models. For the best of our knowledge this is the first attempt to consider the SEP distribution in order to provide a robust framework for quantile regression analysis.

In this paper we propose to use of the SEP distribution to develop a Bayesian robust quantile regression framework. In particular due to the specific characteristics of the SEP distribution we will show how to estimate the quantile function firstly considering the simple linear regression problem then extending it to the generalized additive models (GAM) one. For the latter case we will adopt

the Penalized Spline (P-Spline) approach to carry out the statistical inference. The Bayesian paradigm is implemented by means of a new adaptive Metropolis MCMC sampling scheme with a full set of informative prior. In particular, for the GAM framework, the proposed algorithm turns into an Adaptive Metropolis within Gibbs MCMC for an efficient estimate of the penalization parameter and the P-Spline coefficients.

When dealing with model building the choice of appropriate predictors and consequently the variable selection issue plays an important role. In this paper we approach the problem in the Bayesian quantile regression framework, by considering the Bayesian version of Lasso penalization methodology introduced by Tibshirani (1996). In particular for the linear quantile regression model we will assume, as prior distribution on each regressors, the generalized version of the univariate independent Laplace distribution proposed by Park and Casella (2008) and Hans (2009) already considered in Alhamzawi et al. (2012). With this prior we shrink each parameter separately. As second step, when dealing with the GAM models we generalize the Lang and Brezger (2004) second order random walk prior for the Spline coefficients assuming a multivariate Laplace distribution accounting for a correlation structure among parameters. This prior corresponds to the group lasso penalty one of Yuan and Lin (2006), Meier et al. (2008) and Li et al. (2010) which in the spline contest has a natural interpretation in terms of knots associated with each regressor.

To analyze the performance of the proposed models we consider simulation studies in which we control for the weight of the outliers, the number of the parameters, the shape of the regressors and the presence of heteroschedasticity. Furthermore we analyze three popular real dataset: the corrected version (see Li et al. 2010) of the Boston housing data first analyzed by Harrison and Rubinfeld (1978); the Munich rental dataset with geoaddivitive spatial effect considered in Rue and Held (2005) and Yue and Rue (2011) among the others; the Barro growth data firstly studied by Barro and Sala i-Martin (1995) and then extended in the quantile regression framework by Koenker and Machado (1999). Compared with the existing literature, the models we propose introduce robustness, variable selection and non linearity in the estimation process, providing a more flexible framework and new interesting interpretation of some regression coefficient and, on average, lower posterior standard deviations.

The remainder of the paper is organized as follows. In Section 2, we introduce the SEP distribution and discuss its properties relevant to model conditional quantiles as function of exogenous covariates. In Section 3 we introduce the model specification and the MCMC algorithms proposed. In Section 4 we approach the non-linear extension of the linear quantile approach via GAM models. Section 5 explores the sampling performances of the proposed models through some simulation experiments. Section 6 discusses three well known empirical applications while Section 7 concludes.

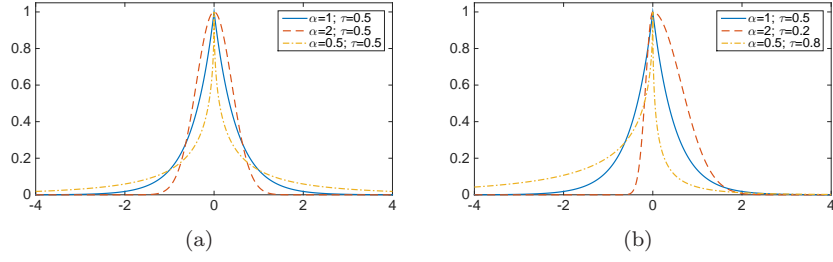


Figure 1: Plot of the SEP distribution for different level of the shape parameter (1(a)) and skewness parameter (1(b)) with $\sigma = 1.0$ and $\mu = 0.0$.

2. The Skewed Exponential Power distribution

Zhu and Zinde-Walsh (2009) have recently proposed a parametrization of the SEP distribution introduced by Fernandez and Steel (1998) particularly convenient when quantiles are the main concern.

Definition 2.1. A random variable $Y \in \mathbb{R}$ is said to be Skewed Exponential Power distributed, i.e. $Y \sim \mathcal{SEP}(\mu, \sigma, \alpha, \tau)$, if its density has the following form:

$$f_{\text{SEP}}(y; \mu, \sigma, \alpha, \tau) = \begin{cases} \frac{1}{\sigma} \kappa_{\text{EP}}(\alpha) \exp \left\{ -\frac{1}{\alpha} \left(\frac{\mu - y}{2\tau\sigma} \right)^\alpha \right\}, & \text{if } y \leq \mu \\ \frac{1}{\sigma} \kappa_{\text{EP}}(\alpha) \exp \left\{ -\frac{1}{\alpha} \left(\frac{y - \mu}{2(1-\tau)\sigma} \right)^\alpha \right\}, & \text{if } y > \mu, \end{cases} \quad (1)$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma \in \mathbb{R}^+$ and $\alpha \in (0, \infty)$ are the scale and shape parameters respectively, while $\kappa_{\text{EP}} = \left[2\alpha^{\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right) \right]^{-1}$ with $\Gamma(\cdot)$ being the complete gamma function. Moreover, the parameter $\tau \in (0, 1)$ controls for the skewness of the distribution.

One of the most nice property of (1), which induces us to propose it for quantile regression inference, is that the location parameter μ coincides with the τ -level quantile (we will theoretically prove it in Appendix A). It can be also shown (see Zhu and Zinde-Walsh, 2009) that the kurtosis of the SEP is directly determined by its parameter α . In Figure 1 we present the pdf of the SEP distribution for different values of shape (α) and skewness (τ) parameters, with fixed values of the location and scale parameters $(\mu, \sigma) = (0, 1)$. It is worth noting that, for a fixed value of $\tau = 0.5$ (see subfigure 1(a)), we retrieve the Laplace and the Normal distribution when the shape parameter is equal to $\alpha = 1$ and $\alpha = 2$, respectively. Moreover the smaller is α , the fatter are the tails of the distribution and in particular as $\alpha \rightarrow 0$ the SEP becomes the Cauchy distribution while as $\alpha \rightarrow \infty$ it becomes equal to the Uniform one. It is hence evident the importance of the parameter α in capturing the behaviour of the tails which may be fundamental when outliers or heavy tails data are modelled. Furthermore, subfigure 1(b) displays the behavior of the SEP for different combination of α and τ . In this

case, the ALD ($\alpha = 1$) and the Skew Normal distribution ($\alpha = 2$) can be obtained because of the role of the skewness parameter τ . In the same figure, it should be also evident the relation between τ and the location parameter μ . For a fixed μ ($\mu = 0$ in the graph) by varying τ , the shape of the distribution changes in such a way that μ becomes its quantile of level τ .

3. Robust Bayesian linear quantile regression

In this section we propose the use of the SEP distribution to implement the Bayesian inference for linear quantile regression combined with the prior distributions specification. Since we are interested in Lasso penalization problem in order to achieve sparsity within the quantile regression model, we propose as prior distribution for the regression parameters, a generalized version of the univariate independent Laplace distribution proposed by Park and Casella (2008) and Hans (2009). In line with Alhamzawi et al. (2012), for each quantile regression parameter we assume a Laplace distribution having different scale parameter in order to shrink each regression parameter in a different way. To achieve the Bayesian procedure we provide an adaptive MCMC sampling scheme obtained by running a block-move Independent Metropolis within Gibbs.

3.1. Model specification

Let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)$ a random sample of T observations and $\mathbf{X}_t = (1, X_{t,1}, \dots, X_{t,p-1})'$, $t = 1, 2, \dots, T$ the associated set of p covariates. Consider the following linear quantile regression model

$$Y_t = \mathbf{X}_t' \boldsymbol{\beta}_\tau + \varepsilon_t, \quad t = 1, 2, \dots, T, \quad (2)$$

where $\boldsymbol{\beta}_\tau = (\beta_{\tau,0}, \beta_{\tau,1}, \dots, \beta_{\tau,p-1})'$ is the vector of p unknown regression parameters varying with the quantile τ level. Here, ε_t , for any $t = 1, 2, \dots, T$, are independent random variables which are supposed to have zero τ -th quantile and constant variance. Assuming $\mathbf{y} = (y_1, y_2, \dots, y_T)$ a realization of \mathbf{Y} and \mathbf{x}_t a realization of \mathbf{X}_t , then the likelihood function for the model (2) based on the SEP distribution (1) with fixed τ can be written as

$$\begin{aligned} \mathcal{L}_\tau(\boldsymbol{\beta}_\tau, \sigma, \alpha, | \mathbf{y}, \mathbf{x}_t) &= \prod_{t=1}^T \frac{1}{2\sigma} \frac{\alpha^{-\frac{1}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})} \left[\exp \left\{ -\frac{1}{\alpha} \left(\frac{\mathbf{x}_t' \boldsymbol{\beta}_\tau - y_t}{2\tau\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t \leq \mathbf{x}_t' \boldsymbol{\beta}_\tau)} \right. \\ &\quad \left. + \exp \left\{ -\frac{1}{\alpha} \left(\frac{y_t - \mathbf{x}_t' \boldsymbol{\beta}_\tau}{2(1-\tau)\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t > \mathbf{x}_t' \boldsymbol{\beta}_\tau)} \right] \\ &= \frac{1}{(2\sigma)^T} \frac{\alpha^{-\frac{T}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})^T} \left[\exp \left\{ -\frac{1}{\alpha} \sum_{t=1}^T \left(\frac{\mathbf{x}_t' \boldsymbol{\beta}_\tau - y_t}{2\tau\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t \leq \mathbf{x}_t' \boldsymbol{\beta}_\tau)} \right. \\ &\quad \left. + \exp \left\{ -\frac{1}{\alpha} \sum_{t=1}^T \left(\frac{y_t - \mathbf{x}_t' \boldsymbol{\beta}_\tau}{2(1-\tau)\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t > \mathbf{x}_t' \boldsymbol{\beta}_\tau)} \right], \quad (3) \end{aligned}$$

where in this case the parameter μ of equation (1) has been replaced by the regression function $\mu = \mathbf{x}_t' \boldsymbol{\beta}_\tau$. As discussed in the previous section, due to the property of the SEP distribution, the regression function $\mathbf{x}_t' \boldsymbol{\beta}_\tau$ corresponds to the conditional τ -level quantile of Y_t i.e. $Q_\tau(Y_t | \mathbf{X}_t = \mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\beta}_\tau$. In what follows, we omit the subscript τ for sake of simplicity.

The Bayesian inferential procedure requires the specification of the prior distribution for the unknown vector of parameters $\boldsymbol{\Xi} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \alpha)$. As mentioned before, for the parameters of the regression function we generalize the prior proposed in Park and Casella assuming the hierarchical structure in (5) and (6) which allows to efficiently shrink each parameter. The priors for the parameters are:

$$\pi(\boldsymbol{\Xi}) = \pi(\boldsymbol{\beta} | \boldsymbol{\gamma}) \pi(\boldsymbol{\gamma}) \pi(\sigma) \pi(\alpha), \quad (4)$$

with

$$\pi(\boldsymbol{\beta} | \boldsymbol{\gamma}) \propto \prod_{j=1}^p L_1(\beta_j | 0, \gamma_j) \quad (5)$$

$$\pi(\boldsymbol{\gamma}) \propto \prod_{j=1}^p \mathcal{G}(\gamma_j | \psi, \varpi) \quad (6)$$

$$\pi(\sigma) \propto \mathcal{IG}(a, b) \quad (7)$$

$$\pi(\alpha) \propto \mathcal{B}(c, d) \mathbb{1}_{(0,2)}(\alpha), \quad (8)$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$, $(\psi, \varpi, a, b, c, d)$ are given positive hyperparameters and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)$ are the parameters of the univariate Laplace distribution:

$$L_1(\beta_j | 0, \gamma_j) = \frac{\gamma_j}{2} \exp\{-\gamma_j |\beta_j|\} \mathbb{1}_{(-\infty, +\infty)}(\beta_j). \quad (9)$$

with zero location and γ_j scale parameter. In (6)-(8) \mathcal{G} , \mathcal{IG} and \mathcal{B} denote the Gamma, Inverse Gamma and Beta distributions, respectively. As known due to its characteristics, the Laplace distribution is the Bayesian counterpart of the Lasso penalization methodology introduced by Tibshirani (1996) to achieve sparsity within the classical regression framework. The original Bayesian Lasso, see also, e.g., Park and Casella (2008) and Hans (2009), introduces the same univariate independent Laplace prior distribution for each regression parameters. Here, as in Alhamzawi et al. (2012), we consider a more general case using the parameters γ_j , $j = 1, 2, \dots, p$ allowing us to overcome the problem that may arise in presence of regressors with different scales of measurement by shrinking each regression parameter in a different way.

As shown in Park and Casella (2008) and Kozumi and Kobayashi (2011), the Laplace distribution can be expressed as a location-scale mixture of Gaussians which adapted to our case becomes

$$L_1(\beta_j | 0, \gamma_j) = \int_0^\infty \frac{1}{\sqrt{2\pi\omega_j}} \exp\left\{-\frac{\beta_j^2}{2\omega_j}\right\} \frac{\gamma_j^2}{2} \exp\left\{-\frac{\gamma_j^2 \omega_j}{2}\right\} d\omega_j, \quad (10)$$

for $j = 1, 2, \dots, p$, where the mixing variable is exponentially distributed with shape parameter $2/\gamma_j^2$. Furthermore, to retain a parsimonious model parameterization, we introduce a second layer hierarchical prior representation for the vector of shape parameters γ , in equation (6). Using the location–scale representation of the Laplace distribution, the prior structure defined in equations (5)–(6), can be represented as follows

$$\beta \mid \omega \sim \mathcal{N}_p(\beta \mid \mathbf{0}_p, \Omega) \quad (11)$$

$$\omega_j \mid \gamma_j \sim \mathcal{E}(\omega_j \mid 2/\gamma_j^2) \quad (12)$$

$$\gamma_j \sim \mathcal{G}(\gamma_j \mid \psi, \varpi), \quad (13)$$

where $\mathbf{0}_p$ is a column vector of zeros of dimension p , $\omega = (\omega_1, \omega_2, \dots, \omega_p)'$, $\Omega = \text{diag}\{\omega_j, j = 1, 2, \dots, p\}$ and \mathcal{E} is the exponential distribution. Concerning the specification of the values for the hyperparameters of the prior distributions, typically, vague priors are imposed on the scale σ because it is regarded as a nuisance parameter, see e.g. Yu and Moyeed (2001) and Tokdar and Kadane (2012). Concerning the prior specification for the shape parameter α , we impose a Beta distribution with $c = 2$ and $d = 2$ in order to allow for a large prior variance without incurring in the problem of U-shaped Beta distribution which gives large probability mass to extreme values. Moreover, we extend the Beta distribution to cover the support $\alpha \in (0, 2)$ where the special case $\alpha = 2$ allows to consider the so called conditional “expectile” of Newey and Powell (1987), while the case $\alpha = 1$ the conditional quantiles based on the ALD introduced by Yu and Moyeed (2001). As mentioned in Section 2, the parameter α regulates the tails–fatness of the SEP distribution so that smaller values implies larger probabilities of extreme observations. Therefore, choosing $\alpha \in (0, 2)$ we encompass both quantile and expectile regression issue addressing at the same time the robustness task relying on a distribution with fatter tails than the Skew Normal.

In the following Section, we introduce the Bayesian parameter estimation procedure which aims to simulate from the posterior distribution using an Adaptive Independent Metropolis–Hastings MCMC algorithm.

3.2. Adaptive IMG for linear quantile regression

The Bayesian inference is carried out using an adaptive MCMC sampling scheme based on the following posterior distribution

$$\pi(\Xi \mid \mathbf{y}, \mathbf{x}) \propto \mathcal{L}_\tau(\beta, \sigma, \alpha \mid \mathbf{y}, \mathbf{x}) \pi(\beta \mid \gamma) \pi(\gamma) \pi(\sigma) \pi(\alpha), \quad (14)$$

where $\mathcal{L}_\tau(\beta, \sigma, \alpha \mid \mathbf{y}, \mathbf{x})$ indicates the likelihood function specified in equation (3). After choosing a set of initial values for the parameter vector $\Xi^{(0)}$, simulations from the posterior distribution at the i –th iteration of $\Xi^{(i)}$, for $i = 1, 2, \dots$, are obtained by running iteratively a block–move Independent Metropolis within Gibbs (IMG). The simulation algorithm requires as first step the specification of a proposal distribution for the parameters (β, σ, α) .

To propose a move for each block of the parameters, we choose the following proposal distributions:

$$q(\boldsymbol{\beta}) \sim \mathcal{N}_p\left(\boldsymbol{\beta} \mid \boldsymbol{\mu}_{\boldsymbol{\beta}}^{(i)}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{(i)}\right) \quad (15)$$

$$q(\sigma) \sim \mathcal{N}_1\left(\tilde{\sigma} \mid \mu_{\tilde{\sigma}}^{(i)}, \psi_{\tilde{\sigma}}^{(i)}\right) \left| \frac{\partial \tilde{\sigma}}{\partial \sigma} \right| \quad (16)$$

$$q(\alpha) \sim \mathcal{N}_1\left(\mu_{\alpha}^{(i)}, \psi_{\alpha}^{(i)}\right) \mathbb{1}_{(0,2)}(\alpha), \quad (17)$$

where the scale parameter $\tilde{\sigma} = \log(\sigma)$ is considered on a log-scale and subsequently transformed to preserve positiveness. The jacobian term in equation (16) is required to get the distribution of the transformation $\sigma = \exp(\tilde{\sigma})$. At each iteration $i = 1, 2, \dots$, the IMG algorithm proceeds by simulating a candidate draw from each parameter block, i.e. $\Upsilon^* = (\xi_1^*, \xi_2^*, \xi_3^*) = (\boldsymbol{\beta}^*, \sigma^*, \alpha^*)$ which is subsequently accepted or rejected. The generic probability that the proposed candidate parameter ξ_j^* , for $j = 1, 2, 3$ becomes the new state of the chain is evaluated on the basis of the following acceptance probability

$$\lambda\left(\xi_j^{(i-1)}, \xi_j^*\right) = \min \left\{ 1, \frac{\mathcal{L}\left(\xi_j^*, \boldsymbol{\Xi}_{-j}^{(i-1)} \mid \mathbf{y}, \mathbf{x}\right)}{\mathcal{L}\left(\xi_j^{(i-1)}, \boldsymbol{\Xi}_{-j}^{(i-1)} \mid \mathbf{y}, \mathbf{x}\right)} \frac{\pi\left(\xi_j^*\right)}{\pi\left(\xi_j^{(i-1)}\right)} \frac{q\left(\xi_j^{(i-1)}\right)}{q\left(\xi_j^*\right)} \right\},$$

for $j = 1, 2, 3$, where $\lambda\left(\xi_j^{(i-1)}, \xi_j^*\right)$ indicates the probability to move from the old to the proposed state of the chain, $\pi(\cdot)$ is the generic prior given in equations (5) - (8) and $\boldsymbol{\Xi}_{-j}^{(i-1)}$ refers to the whole set of parameters at iteration $i - 1$ without the j -th element of Υ^* . To complete the algorithm we sample (ω_j, γ_j) , for $j = 1, 2, \dots, p$ with a Gibbs step by simulating directly from the respective full conditional distributions

$$\begin{aligned} \omega_j \mid \beta_j^{(i)}, \gamma_j^{(i-1)} &\sim \mathcal{GIG}\left(\omega_j \mid \frac{1}{2}, \beta_j^{(i)2}, \gamma_j^{(i-1)2}\right) \\ \gamma_j^{2(i)} \mid \omega_j^{(i)} &\sim \mathcal{G}\left(\gamma_j^2 \mid \psi + 1, \varpi + \frac{\omega_j^{(i)}}{2}\right). \end{aligned}$$

where \mathcal{GIG} denotes the Generalized Inverse Gaussian distribution. Since most of the statistical properties of the Markov chain as well as the performance of the Monte Carlo estimators crucially depend on the definition of the proposal distribution $q(\cdot)$ (see Andrieu and Moulines, 2006 and Andrieu and Thoms, 2008) we improve the basic IMG-MCMC algorithm with an additional step

adapting the proposal parameters using the following equations:

$$\boldsymbol{\mu}_\beta^{(i+1)} = \boldsymbol{\mu}_\beta^{(i)} + \varsigma^{(i+1)} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right), \quad (18)$$

$$\boldsymbol{\Sigma}_\beta^{(i+1)} = \boldsymbol{\Sigma}_\beta^{(i)} + \varsigma^{(i+1)} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right) \left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right)', \quad (19)$$

$$\mu_{\tilde{\sigma}}^{(i+1)} = \mu_{\tilde{\sigma}}^{(i)} + \varsigma^{(i+1)} \left(\tilde{\sigma} - \mu_{\tilde{\sigma}}^{(i)} \right), \quad (20)$$

$$\psi_{\tilde{\sigma}}^{(i+1)} = \psi_{\tilde{\sigma}}^{(i)} + \varsigma^{(i+1)} \left(\tilde{\sigma} - \mu_{\tilde{\sigma}}^{(i)} \right)^2, \quad (21)$$

$$\mu_\alpha^{(i+1)} = \mu_\alpha^{(i)} + \varsigma^{(i+1)} \left(\alpha - \mu_\alpha^{(i)} \right), \quad (22)$$

$$\psi_\alpha^{(i+1)} = \psi_\alpha^{(i)} + \varsigma^{(i+1)} \left(\alpha - \mu_\alpha^{(i)} \right)^2, \quad (23)$$

where $\varsigma^{(i+1)}$ denotes a tuning parameter that should be carefully selected at each iteration to ensure the convergence and the ergodicity of the resulting chain (see Andrieu and Moulines, 2006). Roberts and Rosenthal (2007) provide two conditions for the convergence of the chain: the diminishing adaptation condition, which is satisfied if and only if $\varsigma^{(i)} \rightarrow 0$, as $i \rightarrow +\infty$, and the bounded convergence condition, which essentially guarantees that all transition kernels considered have bounded convergence time. Andrieu and Moulines (2006) show that both conditions are satisfied if and only if $\varsigma^{(i)} \propto i^{-d}$ where $d \in [0.5, 1]$. For those reasons we choose $\varsigma^{(i)} = \frac{1}{Ci^{0.5}}$ where C is set to 10, i.e. $C = 10$. As argued by Roberts and Rosenthal (2007), together these two conditions ensure asymptotic convergence and a weak law of large numbers for this algorithm.

4. Nonlinear extension

In this section, we propose an additive extension of the robust linear quantile regression model considered previously to the class of Generalized Additive Models (GAM) introduced by Hastie and Tibshirani (1986). We will set up GAM models using the SEP likelihood considered before. In order to define the quantile function we make use of the P-Spline functions ending up with a semi-parametric problem. The Bayesian analysis is carried out by generalizing the Lang and Brezger (2004) second order random walk prior for the Spline coefficients assuming a multivariate Laplace distribution accounting for a correlation structure among parameters able to take into account for selection variable issue.

4.1. Non-linear model specification

Generalized Additive models extend multiple linear regression by allowing for the response variable to be modeled as sum of unknown smooth functions of continuous covariates. The aim of this section is to set up a robust non linear and semi-parametric framework for quantile regression following a GAM approach using the SEP likelihood. In particular we assume that the τ -level conditional

quantile can be modeled as a parametric component jointly with a sum of smooth functions as follows:

$$Q_\tau(Y_t | \mathbf{X}_t = \mathbf{x}_t, \mathbf{Z}_t = \mathbf{z}_t) = \mathbf{x}_t' \boldsymbol{\beta} + \sum_{j=1}^J f_j(z_{j,t}), \quad (24)$$

where $\mathbf{x}_t' \boldsymbol{\beta}$ is the parametric component while each $f_j(z_{j,t})$ is a nonparametric continuous smooth function and $\mathbf{z}_t = (z_{1,t}, z_{2,t}, \dots, z_{J,t})'$ is an additional set of covariates. To implement the statistical analysis we assume that the nonparametric component $f_j(z_{tj})$ can be approximated using a polynomial spline of order d , with $k+1$ equally spaced knots between $\min(\mathbf{z}_t)$ and $\max(\mathbf{z}_t)$. Using the well known representation of splines in terms of linear combinations of B-splines, we can rewrite equation (24) as:

$$Q_\tau(Y_t | \mathbf{X}_t = \mathbf{x}_t, \mathbf{Z}_t = \mathbf{z}_t) = \mathbf{x}_t' \boldsymbol{\beta} + \sum_{j=1}^J \sum_{\nu=1}^{k+d} \theta_{j,\nu} B_{j,\nu}(z_{tj}), \quad (25)$$

where $B_{j,\nu}(z_{tj})$ denote B-spline basis functions and $\theta_{j,\nu}$ are the unknown coefficients. In this framework, the value of the estimated coefficients and the shape of the fitted functions strongly depend upon the number and the position of the knots. With respect to the position, in absence of any prior information we consider equidistant knots as a natural choice. Regarding the number of knots, to catch properly the smoothness of the data a careful trade off needs to be considered between few and too many knots since it may cause underfitting or overfitting respectively. A possible solution to this problem is known as Penalized Spline (P-Spline) proposed by O'Sullivan (1986 and 1988) and generalized by Eilers and Marx (1996) which relies on the introduction of a penalty element on the first or second differences of the B-Spline coefficients. This setting has been embedded in the Bayesian framework by Lang and Brezger (2004), Brezger and Lang (2006) and Brezger and Steiner (2008) using a second order random walk for all the B-Spline coefficients, i.e.:

$$\theta_{j,\nu} = 2\theta_{j,\nu-1} - \theta_{j,\nu-2} + u_{j,\nu}, \quad \forall j = 1, 2, \dots, J, \quad \forall \nu = 1, 2, \dots, k+d, \quad (26)$$

where the generic stochastic component $u_{j,\nu} \sim \mathcal{N}(0, h_j)$ and $\theta_{j,1}$ and $\theta_{j,2}$ are initialized with diffuse priors, i.e., $\pi(\theta_{j,\nu}) \propto 1$, for $\nu = 1, 2$. In their work Lang and Brezger (2004) assume that the stochastic components $u_{j,\nu}$ driving the random walk process are independent, i.e. $u_{j,\nu} \perp u_{k,\nu}$, for all $j, k = 1, 2, \dots, J$ with $j \neq k$. Since there are no reasons to assume a priori $u_{j,\nu}$ and $u_{k,\nu}$ independent ($\forall j, k$) here we consider an extension of (26) and we assume a multivariate Laplace distribution on the vector of regressors accounting for a correlation structure among them. Moreover it can easily be proved, that the original marginal shrinkage effect is preserved under the assumed prior structure, because each marginal prior reduces to a univariate Laplace, see, e.g., Kotz et al. (2001). Moreover using the Laplace distribution as prior distribution allows to extend the Bayesian Lasso approach to estimation.

Let $\mathbf{u}_j = (u_{j,1}, u_{j,2}, \dots, u_{j,k+d})$, we assume $\mathbf{u}_j \sim \mathcal{AL}_{k+d}(0, \mathbf{I}_{k+d})$, where \mathcal{AL}_{k+d} denotes the multivariate Laplace distribution and \mathbf{I}_{k+d} is the identity matrix of dimension $k+d$. Furthermore, let \mathbf{D}_δ to be the difference matrix of dimension $(k+d-\delta) \times (k+d)$ and $\delta = 2$ is the order of the differential operator, such that $\mathbf{D}_\delta \boldsymbol{\theta}_j = \mathbf{u}_j$, then

$$\pi(\boldsymbol{\theta}_j \mid h_j) \sim \mathcal{AL}_{k+d}(0, h_j(\mathbf{D}'_\delta \mathbf{D}_\delta)), \quad \forall j = 1, 2, \dots, J, \quad (27)$$

having density

$$\pi(\boldsymbol{\theta}_j \mid h_j) = C |\mathbf{D}'_\delta \mathbf{D}_\delta|^{\frac{1}{2}} h_j^{d+k} \exp \left\{ -h_j [\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j]^{\frac{1}{2}} \right\}, \quad (28)$$

where $C = \frac{\sqrt{2\pi}}{\Gamma(\frac{d+k+1}{2})}$. As shows in Kotz et al. (2001), the multivariate Laplace distribution can be expressed as a location-scale mixture of Gaussians, where the mixing variable follows a Gamma distribution

$$\pi(\boldsymbol{\theta}_j \mid \phi_j) \sim \mathcal{N}_{k+d} \left(0, \phi_j (\mathbf{D}'_\delta \mathbf{D}_\delta)^{-1} \right) \quad (29)$$

$$\pi(\phi_j \mid h_j) \sim \mathcal{G} \left(\frac{k+d+1}{2}, \frac{h_j^2}{2} \right), \quad (30)$$

for $j = 1, 2, \dots, J$.

It is easy to show how to retrieve (28) from (29) - (30) by integrating out the augmented variable ϕ_j , i.e.

$$\begin{aligned} \pi(\boldsymbol{\theta}_j \mid h_j) &= \int_{\mathbb{R}} \frac{|\mathbf{D}'_\delta \mathbf{D}_\delta|^{\frac{1}{2}} \exp \left\{ -\frac{\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j}{2\phi_j} \right\} \left(\frac{h_j^2}{2} \right)^{\frac{d+k+1}{2}} \phi_j^{\frac{d+k-1}{2}}}{(2\pi\phi_j)^{\frac{k+d}{2}} \Gamma\left(\frac{d+k+1}{2}\right)} \\ &\quad \times \exp \left\{ \frac{-h_j^2 \phi_j}{2} \right\} d\phi_j \\ &= \frac{\left(\frac{h_j^2}{2} \right)^{\frac{d+k+1}{2}} |\mathbf{D}'_\delta \mathbf{D}_\delta|^{\frac{1}{2}}}{(2\pi)^{\frac{k+d}{2}} \Gamma\left(\frac{d+k+1}{2}\right)} \\ &\quad \times \int_{\mathbb{R}} \phi_j^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j}{\phi_j} + h_j^2 \phi_j \right] \right\} d\phi_j, \quad (31) \end{aligned}$$

where the integrand in the previous equation (31) is proportional to a Generalized Inverse Gaussian distribution $\mathcal{GIG}(p, a, b)$ with parameters $p = \frac{1}{2}$, $a = h_j^2$ and $b = \boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j$ from which we have

$$\int_{\mathbb{R}} \phi_j^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j}{\phi_j} + h_j^2 \phi_j \right] \right\} d\phi_j = \frac{2K_{\frac{1}{2}} \left(\sqrt{h_j^2 (\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j)} \right)}{\left(\frac{h_j^2}{\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j} \right)^{\frac{1}{4}}},$$

where $K_{\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2z}} \exp\{-z\}$. Substituting this latter expression into equation (31) we obtain

$$\begin{aligned} \pi(\boldsymbol{\theta}_j | h_j) &= \frac{\left(\frac{h_j^2}{2}\right)^{\frac{d+k+1}{2}} |\mathbf{D}'_{\delta} \mathbf{D}_{\delta}|^{\frac{1}{2}} \frac{\sqrt{2\pi} \exp\left\{-h_j \sqrt{\boldsymbol{\theta}'_j (\mathbf{D}'_{\delta} \mathbf{D}_{\delta}) \boldsymbol{\theta}_j}\right\}}{[h_j^2 (\boldsymbol{\theta}'_j (\mathbf{D}'_{\delta} \mathbf{D}_{\delta}) \boldsymbol{\theta}_j)]^{\frac{1}{4}}}}{(2\pi)^{\frac{k+d}{2}} \Gamma\left(\frac{d+k+1}{2}\right) \left(\frac{h_j^2}{\boldsymbol{\theta}'_j (\mathbf{D}'_{\delta} \mathbf{D}_{\delta}) \boldsymbol{\theta}_j}\right)^{\frac{1}{4}}} \\ &= \frac{\sqrt{\pi} |\mathbf{D}'_{\delta} \mathbf{D}_{\delta}|^{\frac{1}{2}} h_j^{d+k}}{\Gamma\left(\frac{d+k+1}{2}\right)} \exp\left\{-h_j \sqrt{\boldsymbol{\theta}'_j (\mathbf{D}'_{\delta} \mathbf{D}_{\delta}) \boldsymbol{\theta}_j}\right\}, \end{aligned} \quad (32)$$

which corresponds to the ALD defined in equations (27)–(28). The proposed prior distribution for $\boldsymbol{\theta}_j$ corresponds to the group lasso penalty of Yuan and Lin (2006), Meier et al. (2008) and Li et al. (2010) which accounts for the group structure when performing variable selection. It is worth emphasizing that, in our context, the group variables have a natural interpretation because they correspond to knots accounting for the smoothness level of the same regressor over different regions of the support. The overall smoothness of the fitted curve is controlled by the variance of the error term h_j which correspond to the inverse of the penalization parameter used by Eilers et al. (1996) in the frequentist framework. We choose a conjugate Gamma prior for h_j^2 , that is $h_j^2 \sim \mathcal{G}(a^{(h)}, b^{(h)})$ with $a^{(h)} = b^{(h)} = 0.001$. Different choice of hyper parameters may be considered but they all bring to very similar results. Summarizing, putting a Gamma prior on h_j^2 , and assuming the prior structure defined in equations (5)–(6) for the shape and scale parameters (σ, α) , we then have the following hierarchical model

$$y_t = x'_t \boldsymbol{\beta} + \sum_{j=1}^J \mathbf{B}_j^z \boldsymbol{\theta}_j + \epsilon_t, \quad \epsilon_t \sim \mathcal{SE}\mathcal{P}(0, \tau, \sigma, \alpha) \quad (33)$$

$$\boldsymbol{\beta} | \boldsymbol{\omega} \sim \mathcal{N}_p(\boldsymbol{\beta} | \mathbf{0}_p, \boldsymbol{\Omega}) \quad (34)$$

$$\omega_k | \gamma_k \sim \mathcal{E}(\omega_k | 2/\gamma_k^2) \quad (35)$$

$$\gamma_k \sim \mathcal{G}(\gamma_k | \psi, \varpi), \quad \forall k = 1, 2, \dots, p \quad (36)$$

$$\begin{aligned} \boldsymbol{\theta}_j | \phi_j &\sim \mathcal{N}_{k+d}\left(0, \phi_j (\mathbf{D}'_{\delta} \mathbf{D}_{\delta})^{-1}\right) \\ \phi_j | h_j &\sim \mathcal{G}\left(\frac{d+k+1}{2}, \frac{h_j^2}{2}\right) \end{aligned} \quad (37)$$

$$h_j^2 \sim \mathcal{IG}\left(\frac{a^{(h)}}{2}, \frac{b^{(h)}}{2}\right) \quad \forall j = 1, 2, \dots, J, \quad (38)$$

where $\mathbf{B}_j^z = (B_{j,1}(z_{tj}), \dots, B_{j,k+d}(z_{tj}))$.

4.2. Adaptive IMG for quantiles GAM

In order to perform the Bayesian inference the Adaptive MCMC algorithm proposed in Section 3.2 will be slightly modified to deal with the simulation

from the posterior distribution of the generalized quantiles GAM parameters. The posterior distribution becomes equal to

$$\pi(\Xi | \mathbf{y}, \mathbf{x}, \mathbf{z}) \propto \mathcal{L}_\tau(\beta, \sigma, \alpha, \boldsymbol{\vartheta} | \mathbf{y}, \mathbf{x}, \mathbf{z}) \pi(\beta | \gamma) \pi(\gamma) \times \pi(\boldsymbol{\vartheta} | \phi) \pi(\phi | \mathbf{h}) \pi(\mathbf{h}) \pi(\sigma, \alpha) \quad (39)$$

where the vector Ξ contains now three more additional set of parameters, namely $\boldsymbol{\vartheta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_J)$, $\phi = (\phi_1, \phi_2, \dots, \phi_J)$ and $\mathbf{h} = (h_1^2, h_2^2, \dots, h_J^2)$. The likelihood function $\mathcal{L}_\tau(\beta, \sigma, \alpha, \boldsymbol{\vartheta} | \mathbf{y}, \mathbf{x}, \mathbf{z})$ defined in equation (3) for the linear model should be adapted to account for the additional splines coefficients. Here to perform the Bayesian analysis it is necessary to add three more steps to the algorithm described in Section 3.2. Specifically, after having updated all the parameters of the linear part of the model, a candidate for $\boldsymbol{\theta}_j$, for $j = 1, 2, \dots, J$ is drawn from a Gaussian proposal distribution, i.e., $q(\boldsymbol{\theta}_{j,i-1}, \boldsymbol{\theta}_j^*) \sim \mathcal{N}_{k+d}(\boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_j}^{(i)})$ and accepted on the basis of the following acceptance probability

$$\lambda(\boldsymbol{\theta}_j^{(i-1)}, \boldsymbol{\theta}_j^*) = \min \left\{ 1, \frac{\mathcal{L}(\beta^{(i-1)}, \sigma^{(i-1)}, \alpha^{(i-1)}, \boldsymbol{\theta}_j^*, \boldsymbol{\vartheta}_{-j}^{(i-1)} | \mathbf{y}, \mathbf{x}, \mathbf{z})}{\mathcal{L}(\beta^{(i-1)}, \sigma^{(i-1)}, \alpha^{(i-1)}, \boldsymbol{\vartheta}_j^{(i-1)} | \mathbf{y}, \mathbf{x}, \mathbf{z})} \times \frac{\pi(\boldsymbol{\theta}_j^*)}{\pi(\boldsymbol{\theta}_j^{(i-1)})} \frac{q(\boldsymbol{\theta}_j^{(i-1)})}{q(\boldsymbol{\theta}_j^*)} \right\},$$

where $\boldsymbol{\vartheta}_{-j}$ denote the whole set of B-Spline coefficients without the j -th component. Furthermore, as specified in Section 3.2 for the regression parameters, an adaptive step for the mean and the variance of the proposal distribution of each $\boldsymbol{\theta}_j$ is implemented using the following equation

$$\boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i+1)} = \boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)} + \varsigma^{(i+1)} (\boldsymbol{\theta}_j - \boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)}), \quad (40)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}_j}^{(i+1)} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}_j}^{(i)} + \varsigma^{(i+1)} (\boldsymbol{\theta}_j - \boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)}) (\boldsymbol{\theta}_j - \boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)})', \quad (41)$$

for $j = 1, 2, \dots, J$, where ς is the vanishing factor fixed as discussed above. The hyperparameters (ϕ, \mathbf{h}) are updated by single-move Gibbs sampling steps by simulating from the following full conditionals which are proportional to the GIG distribution

$$\begin{aligned} \phi_j | \boldsymbol{\theta}_j^{(i)}, h_j^{(i-1)} &\sim \mathcal{GIG} \left(\phi_j \middle| \frac{1}{2}, h_j^{(i-1)2}, \boldsymbol{\theta}_j' (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j \right) \\ h_j^2 | \phi_j^{(i)} &\sim \mathcal{GIG} \left(h_j^2 \middle| -\frac{a^h}{2}, \phi_j^{(i)}, \frac{b^h}{2} \right). \end{aligned}$$

5. Simulation Studies

In this Section, simulation studies are performed to highlight the improvements in robustness obtained by implementing SEP based quantile regression compared with those obtained by the traditional Bayesian quantile regression based

on the ADL distribution. More specifically, our purpose is to illustrate how the SEP misspecified model assumption in the quantile regression framework generate posterior distributions of the regression parameters centered on the true values. The first simulation experiment assesses the robustness properties of the proposed methodology for quantile estimation when the joint distribution of the couple (Y_i, \mathbf{X}_i) , for $i = 1, 2, \dots, T$, is contaminated by the presence of outliers. The second study shows the effectiveness of shrinkage effect obtained by imposing the Lasso-type prior proposed when multiple quantile linear model is the main concern. The last experiment aims at highlighting the ability of the model to adapt to non-linear shapes when data come from heterogeneous fat-tailed distributions. The hyperparameters of the priors distribution are all chosen such that the priors are non informative. In particular regarding the nuisance parameter σ we choose $a = b = 0.001$ which corresponds to a proper Inverse Gamma distributions with infinite second moments. When lasso prior is assumed the hyperparameters (ψ, ϖ) in the Gamma priors for γ_j are chosen to be 0.1.

5.1. Simple linear quantile regression

For this experiment we consider a sample of $T = 100$ drawn from the following homoskedastic mixture of distributions

$$f((Y, X)', \boldsymbol{\eta}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_L, \boldsymbol{\Sigma}) = \sum_{l=1}^L \eta_l \varphi(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}), \quad (42)$$

where φ denote the density function of a Gaussian distribution with mean $\boldsymbol{\mu}$ and variance and covariance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{\eta} = (\eta_1, \dots, \eta_L)$ is the vector of weights. We fix the number of components equal to $L = 3$, with mixture weights $\boldsymbol{\eta} = (0.85, 0.0725, 0.0725)$, locations and scale matrix specified as

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_3 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}. \quad (43)$$

The quantile regression model implemented is a simple model with only one exogenous variable i.e. $Y_t = X_t \beta + \varepsilon_t$ for $t = 1, 2, \dots, T$. The aim of this toy example is to show the performances of the Bayesian quantile linear regression analysis assuming both the ALD and the SEP likelihood when the data are strongly contaminated by the presence of outliers. Since we have only one regressor, for this illustrative example we use a simplified version of the sampler proposed in Section 3.2, in which a simple Gaussian prior is considered for β . For $\tau = (0.1, 0.5, 0.9)$ we run the MCMC algorithm with $N = 50,000$ iterations and a burn-in of $M = 10,000$. For both the ALD and the SEP distribution assumption, initial values for the parameters to be estimated, namely (β, σ) , are randomly drawn from $\mathcal{N}(0, 1)$ and $\mathcal{IG}(0.001, 0.001)$, respectively. The additional initial value for the parameter α , required only for the SEP distribution case, is randomly drawn from $\mathcal{B}(2, 2)$.

Figure 2 depicts the estimated regression lines as well as the 95% HPD credible sets. The blue line refers to the ALD estimation while the red line to the

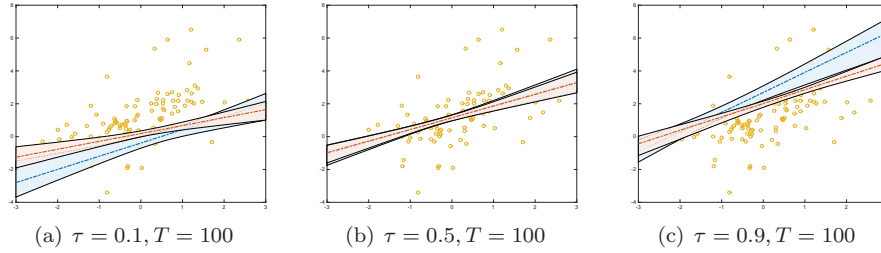


Figure 2: Contaminated data example. Comparison between Bayesian quantile regression based on the ALD (blue) and the SEP distribution (red) for different values of the quantile confidence level $\tau = (0.1, 0.5, 0.9)$ and sample sizes $T = 100$. Shaded areas denote 95% HPD credible sets.

SEP one. It can be easily observed that the two curves overlap for $\tau = 0.5$ and increasingly diverge for more extreme level quantiles i.e. for $\tau = (0.1, 0.9)$. It is in fact the case that for $\tau = 0.5$ the posterior mean of α is very close to one, which implies that the SEP reduces to the the ALD distribution. As far as we move

Parameter	ALD			SEP		
	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
β_0	-0.391 (0.176)	1.149 (0.093)	2.688 (0.237)	0.186 (0.089)	1.144 (0.086)	2.011 (0.100)
β_1	0.801 (0.151)	0.735 (0.106)	1.207 (0.144)	0.428 (0.094)	0.709 (0.093)	0.825 (0.074)
σ	3.844 (0.386)	2.105 (0.212)	3.478 (0.349)	1.049 (0.153)	0.862 (0.112)	0.989 (0.150)
α	-	-	-	0.596 (0.094)	0.832 (0.142)	0.504 (0.068)

Table 1: Contaminated data example. Estimated parameters for different levels of the quantile confidence level $\tau = (0.1, 0.5, 0.9)$ and $T = 100$. Standard deviations in parenthesis.

away from the median, it is notable the differences in the estimated regression quantile parameters under the ALD and the SEP assumption. Looking at the subplots 2(a) and 2(c) it is evident that the intercept and the slope of the regression line obtained using ALD distribution is strongly influenced by the 7.25% of the outliers from the two external components of the mixture. In those cases, the estimated α of the SEP is considerably smaller than one. The estimation of the β 's parameters is therefore made under a distribution with fatter tails than the ALD strongly penalizing more extreme observations and providing, as consequence, more robust results.

For the regression parameters Table 1 contains the estimated posterior means and standard deviations under the ALD and the SEP assumption. Under the data generating process the theoretical slope should be always equal to 0.6. It can be seen that moving from the median to more extreme quantiles, the

posterior mean of the intercept and the slope, estimated with the ALD, is farther from the true value than those obtained with the SEP. In addition it is worth noting that also the standard errors are always lower implying that the estimated are more sharp when using the SEP distribution .

5.2. Multiple quantile regression

In this section, we carry out a Monte Carlo simulation study specifically tailored to evaluate the performance of the model when the Lasso prior (5) is considered for the regression parameters. The simulation are similar to the one proposed in Li et al. (2010) and Alhamzawi et al. (2012). Specifically, we simulate $T = 200$ observations from the linear model $Y_t = \mathbf{X}_t' \boldsymbol{\beta} + \varepsilon_t$, where the true values for the regressors are set as follows:

- Simulation 1. $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)'$,
- Simulation 2. $\boldsymbol{\beta} = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)'$,
- Simulation 3. $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0)'$,

The first simulation corresponds to a sparse regression case, the second to a dense case while the third to a very sparse one. The covariates are independently generated from a $\mathcal{N}(0, \Sigma)$ with $\sigma_{i,j} = 0.5^{|i-j|}$. Two different choices for the error terms distribution of the generating process distribution are considered for each simulation study. The first choice is a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, with μ chosen so that the τ -th quantile is 0, while σ^2 is set as 9, as in Li et al. (2010). The second choice is a Generalized Student't distribution $\mathcal{GS}(\mu, \sigma^2, \nu)$ with two degrees of freedom, i.e. $\nu = 2$, $\sigma^2 = 9$ and μ chosen so that the τ -th quantile is 0. For three different quantile level, $\tau = (0.10, 0.5, 0.9)$ we run 50 simulations for each vector of parameters ($\boldsymbol{\beta}$) and each choice of the error term. Table 2 reports the median of mean absolute deviation (MMAD), i.e. $\text{median}\left(\frac{1}{200} \sum_{t=1}^{200} |x_t' \hat{\boldsymbol{\beta}} - x_t' \boldsymbol{\beta}|\right)$, and the median of the parameters $\hat{\boldsymbol{\beta}}$ over 50 estimates. To be concise only results for simulation1 are reported since results from the other two simulations are similar. It is immediate to see that the proposed Bayesian quantile regression method based on the SEP likelihood performs better in terms of MMAD for both the distributions of the error term. This evidence confirms that the presence of the shape parameter α in the likelihood allows to better capture the behavior of the data. The estimated shape parameter is indeed greater and lower then 1 in the Gaussian and Generalized Student case respectively, giving us a more reliable estimation of the vector $\boldsymbol{\beta}$ regardless to the tails weight of the distribution of the error term. These results are confirmed in simulation 2 and simulation 3 (not reported here) in which we exasperate the density and the sparsity in the structure of the predictors. Furthermore, looking at the values of $\boldsymbol{\beta}$ we can see that the proposed robust method reduces the bias of the estimates for all the quantile confidence levels. Concerning the shrinkage ability of the proposed estimator we observe that where the true parameters are zero, the SEP distribution is able to identify them better than the ALD.

Error distribution	Par.	ALD			SEP		
		$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.90$
Gaussian	MMAD	1.0131	1.1008	1.0579	0.9096	1.0955	0.9708
	β_1	3.1323	3.2209	3.2145	3.0744	3.0036	3.2127
	β_2	1.6408	1.4786	1.6165	1.7656	1.4833	1.6800
	β_3	0.0444	0.0294	0.0267	0.0428	0.0228	0.0186
	β_4	0.0453	0.0243	0.0235	0.0248	0.0191	0.0156
	β_5	1.2731	1.2379	1.3471	1.3969	1.8405	1.4702
	β_6	0.0185	0.0161	0.0205	0.0124	0.0127	0.0128
	β_7	0.0112	0.0106	0.0120	0.0067	0.0063	0.0095
	β_8	0.0073	0.0078	0.0064	0.0038	0.0047	0.0051
Generalized Student t	MMAD	0.5163	0.1807	0.4685	0.4777	0.1789	0.4275
	β_1	3.0630	2.9884	2.9874	3.0826	2.9877	2.9934
	β_2	1.0484	1.3700	1.1366	1.0952	1.3951	1.2110
	β_3	0.0304	0.0144	0.0325	0.0252	0.0135	0.0412
	β_4	0.0258	0.0181	0.0162	0.0263	0.0163	0.0138
	β_5	1.7012	1.9036	1.7701	1.7558	1.9111	1.8052
	β_6	0.0128	0.0085	0.0137	0.0074	0.0072	0.0136
	β_7	0.0055	0.0057	0.0101	0.0052	0.0066	0.0082
	β_8	0.0067	0.0009	0.0002	0.0051	0.0011	-0.0021

Table 2: Multiple regression simulated data example 1. MMADs and estimated parameters for Simulation 1 under the SEP and ALD assumption for the quantile error term.

5.3. Non Linear Model

In this simulation example we illustrate the performances of model assumptions when a simple GAM model is considered with a single continuous smooth non-linear function and where the parametric linear components are set to zero. Following Chen and Yu (2009), we consider two data generating process $y_t = f_j(x_t) + \epsilon_t$, for $t = 1, 2, \dots, T$ and $j = 1, 2$ where f_1 represents the wave function and f_2 the doppler function, defined as follows

$$f_1(x) = 4(x - 0.5) + 2 \exp\left(-256(x - 0.5)^2\right) \mathbb{I}_{(0,1)}(x) \quad (44)$$

$$f_2(x) = (0.2x(1 - 0.2x))^{\frac{1}{2}} \sin\left(\frac{2\pi(1 + \gamma)}{0.2x + \gamma}\right) \mathbb{I}_{(0,1)}(x), \quad (45)$$

with $\gamma = 0.15$. These functions are usually used (see also Denison et al. 1998) to check the nonlinear fitting ability of a proposed model. Starting from these two curves, we generate a sample of $T = 200$ and $T = 512$ observations for the

wave and the doppler functions, respectively using four different sources of error

$$\textbf{Gaussian noise}, \quad \epsilon_t \sim \mathcal{N}(0, 1) \quad (46)$$

$$y_t = f_1(x_t) + \sigma_1 \epsilon_t,$$

$$y_t = f_2(x_t) + \sigma_2 \epsilon_t,$$

$$\textbf{Student-t noise}, \quad \epsilon_t \sim \mathcal{T}_\nu(0, 1), \quad (47)$$

$$y_t = f_1(x_t) + \sigma_1 \epsilon_t,$$

$$y_t = f_2(x_t) + \sigma_2 \epsilon_t,$$

$$\textbf{Linear heterogeneity}, \quad \epsilon_t \sim \mathcal{T}_\nu(0, 1), \quad (48)$$

$$y_t = f(x_t) + \sigma_1(1 + x) \epsilon_t,$$

$$y_t = f(x_t) + \sigma_2(1 + x) \epsilon_t,$$

$$\textbf{Quadratic heterogeneity}, \quad \epsilon_t \sim \mathcal{T}_\nu(0, 1), \quad (49)$$

$$y_t = f_1(x_t) + \sigma_1(1 + x_t^2) \epsilon_t,$$

$$y_t = f_2(x_t) + \sigma_2(1 + x_t^2) \epsilon_t,$$

where $\sigma_1 = \sqrt{0.4}$, $\sigma_2 = \sqrt{0.1}$ and $\nu = 2$. All the considered model specifications are estimated using penalized P-Splines of order 4 imposing a relative large number of equally spaced knots and with a penalization parameter $\delta = 2$, as suggested by Eilers et al. (1996). In particular, we use 20 knots for the wave function and 25 knots for the doppler one because of the presence of many change points. The sampling process is performed using 10,000 iterations with the first 5,000 as burn-in. Table 3 shows the average and the standard errors, for 50 repeats, of the mean squared errors (*mse*) of three different quantile levels for all the described curves. It can be noted that the SEP outperforms almost uniformly the ALD in terms of *mse*. The difference between the two curves is less evident in presence of Gaussian errors, where the ALD shows also a smaller *mse* for the extreme quantiles of the wave function. The improvement in terms of estimation bias becomes greater looking at more heavy tailed and heteroskedastic error distributions. Concerning the wave function the SEP shows a *mse* that is equal to half of that obtained with the ALD at the extreme quantiles. The same conclusions can be drawn for the doppler function that is generally better estimated than the wave.

6. Empirical applications

Three empirical datasets are analyzed in this section: Boston Housing, Munich Rent and Barro growth data. The first dataset is characterized by the presence of many regressors that allows us to emphasize the usefulness of introducing a lasso prior for the regression parameters. The second one also has a large set of regressors but some of them are characterized by a non linear relation with the response variable. For this dataset we highlight that the assumption of a lasso prior within a robust quantile GAM framework leads us to a more precise estimation process. Finally we propose the use of our robust quantile lasso GAM

Model	Noise	Wave			Doppler		
		$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
ALD	Gaussian	0.0054	0.0022	0.0039	0.0002	0.0000	0.0002
		(0.0171)	(0.0070)	(0.0124)	(0.0007)	(0.0002)	(0.0005)
	Student-t	0.0504	0.0034	0.0177	0.0009	0.0001	0.0177
		(0.1593)	(0.0108)	(0.0561)	(0.0027)	(0.0045)	(0.0015)
	Lin. Het.	0.1035	0.0054	0.0627	0.0059	0.0002	0.0039
		(0.3273)	(0.0170)	(0.1979)	(0.0180)	(0.0010)	(0.0124)
	Quad. Het.	0.0505	0.0067	0.0752	0.0018	0.0001	0.0050
		(0.1598)	(0.0210)	(0.2377)	(0.0058)	(0.0006)	(0.0160)
SEP	Gaussian	0.0071	0.0020	0.0078	0.0002	0.0000	0.0005
		(0.0226)	(0.0063)	(0.0248)	(0.0006)	(0.0002)	(0.0016)
	Student-t	0.0251	0.0037	0.0132	0.0007	0.0001	0.0006
		(0.0795)	(0.0117)	(0.0417)	(0.0022)	(0.0045)	(0.0019)
	Lin. Het.	0.0986	0.0046	0.0678	0.0008	0.0003	0.0006
		(0.3118)	(0.0145)	(0.2144)	(0.0026)	(0.0010)	(0.0020)
	Quad. Het.	0.0234	0.0057	0.0305	0.0011	0.0002	0.0008
		(0.0741)	(0.0180)	(0.0965)	(0.0035)	(0.0006)	(0.0026)

Table 3: Non-linear regression simulated data example. MSE of the fitted curves with four sources of noise evaluated over the 200 synthetic replications. Standard deviations in parenthesis.

model to study the Barro growth data by assuming a non linear representation for some regressors. We find a new interesting interpretation of regression parameters while the neoclassical convergence hypothesis is maintained.

6.1. Boston housing data

In this section we analyze the Boston Housing data first considered by Harrison and Rubinfeld (1978) studying the influence of pollution on house prices. In particular in this paper we consider the corrected data of Li *et al.* (2010). The model is based on the log-transformed corrected median values of owner-occupied housing (values in USD 1000) as dependent variable while several exogenous variables are taken into account: the point longitudes and latitudes in decimal degrees (LON and LAT respectively), the per capita crime (CRIM), the proportions of residential land zoned and non-retail business acres per town (ZN and INDUS respectively), a dummy equal to 1 if tract borders Charles River (CHAS), the nitric oxides concentration (NOX), the average numbers of rooms per dwelling (RM), the proportions of owner-occupied units built prior to 1940 (AGE), the weighted distances to five Boston employment centers (DIS), the index of accessibility to radial highways per town (RAD), the full-value property-tax rate per town (TAX), the pupil-teacher ratios per town (PTRATIO), the transformed Black population proportion (B) and percentage values of lower status population (LSTAT) as regressors. To provide a complete description of the conditional distribution of the response variable we consider five different choices of τ , i.e. 0.10 0.25 0.50 0.75 0.90. Moreover in order to show the opportunity of assuming a Lasso prior for the regressor parameters and to show its performances we consider also a Gaussian prior distribution. Results are showed in Table 4 where it is evident that independently of the choice of the prior

Variable	Gaussian Prior					Lasso Prior				
	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
LON	-0.0614 (0.0364)	-0.0297 (0.0416)	-0.0203 (0.0450)	-0.0114 (0.0555)	0.0072 (0.0434)	-0.0287 (0.0166)	-0.0213 (0.0172)	-0.0258 (0.0187)	-0.0261 (0.0176)	-0.0161 (0.0157)
LAT	-0.0250 (0.0582)	0.0251 (0.0678)	0.0386 (0.0732)	0.0531 (0.0937)	0.0816 (0.0729)	0.0103 (0.0276)	0.0221 (0.0290)	0.0168 (0.0311)	0.0121 (0.0296)	0.0238 (0.0262)
CRIM	-0.0230 (0.0032)	-0.0177 (0.0027)	-0.0093 (0.0015)	-0.0063 (0.0016)	-0.0058 (0.0021)	-0.0241 (0.0028)	-0.0178 (0.0027)	-0.0093 (0.0014)	-0.0059 (0.0017)	-0.0032 (0.0014)
ZN	0.0000 (0.0003)	0.0005 (0.0003)	0.0008 (0.0004)	0.0012 (0.0004)	0.0012 (0.0003)	-0.0000 (0.0003)	0.0006 (0.0003)	0.0009 (0.0004)	0.0010 (0.0003)	0.0009 (0.0002)
INDUS	0.0030 (0.0016)	0.0023 (0.0014)	0.0027 (0.0016)	0.0014 (0.0017)	-0.0013 (0.0013)	0.0016 (0.0015)	0.0017 (0.0014)	0.0016 (0.0016)	0.0010 (0.0017)	-0.0027 (0.0011)
CHAS	0.0482 (0.0264)	0.0464 (0.0197)	0.0597 (0.0264)	0.0737 (0.0308)	0.1134 (0.0302)	0.0183 (0.0190)	0.0377 (0.0205)	0.0411 (0.0236)	0.0448 (0.0275)	0.0834 (0.0312)
NOX	-0.3667 (0.1324)	-0.2642 (0.0946)	-0.3672 (0.1119)	-0.4465 (0.1424)	-0.3204 (0.1191)	-0.0397 (0.0463)	-0.0604 (0.0575)	-0.0480 (0.0551)	-0.0416 (0.0524)	-0.0222 (0.0395)
RM	0.2050 (0.0258)	0.2259 (0.0141)	0.2139 (0.0175)	0.2007 (0.0255)	0.2067 (0.0154)	0.2318 (0.0157)	0.2299 (0.0131)	0.2129 (0.0173)	0.2262 (0.0201)	0.2347 (0.0142)
AGE	-0.0013 (0.0005)	-0.0016 (0.0003)	-0.0009 (0.0004)	-0.0000 (0.0006)	0.0006 (0.0003)	-0.0019 (0.0003)	-0.0018 (0.0003)	-0.0011 (0.0004)	-0.0008 (0.0005)	0.0002 (0.0003)
DIS	-0.0337 (0.0063)	-0.0342 (0.0054)	-0.0330 (0.0062)	-0.0337 (0.0063)	-0.0313 (0.0043)	-0.0279 (0.0050)	-0.0303 (0.0047)	-0.0268 (0.0061)	-0.0267 (0.0054)	-0.0257 (0.0033)
RAD	0.0113 (0.0025)	0.0100 (0.0018)	0.0074 (0.0024)	0.0106 (0.0022)	0.0118 (0.0019)	0.0103 (0.0022)	0.0089 (0.0016)	0.0077 (0.0027)	0.0095 (0.0021)	0.0099 (0.0014)
TAX	-0.0007 (0.0001)	-0.0006 (0.0001)	-0.0005 (0.0001)	-0.0004 (0.0001)	-0.0004 (0.0001)	-0.0007 (0.0001)	-0.0006 (0.0001)	-0.0005 (0.0001)	-0.0005 (0.0001)	-0.0004 (0.0001)
PRATIO	-0.0295 (0.0037)	-0.0292 (0.0032)	-0.0318 (0.0039)	-0.0302 (0.0047)	-0.0253 (0.0033)	-0.0300 (0.0024)	-0.0270 (0.0027)	-0.0280 (0.0037)	-0.0245 (0.0035)	-0.0199 (0.0027)
B	0.0006 (0.0001)	0.0006 (0.0001)	0.0007 (0.0001)	0.0007 (0.0001)	0.0006 (0.0002)	0.0006 (0.0001)	0.0006 (0.0001)	0.0007 (0.0001)	0.0008 (0.0001)	0.0009 (0.0001)
LSTAT	-0.0186 (0.0027)	-0.0172 (0.0019)	-0.0189 (0.0023)	-0.0195 (0.0028)	-0.0191 (0.0015)	-0.0161 (0.0018)	-0.0173 (0.0019)	-0.0205 (0.0023)	-0.0179 (0.0025)	-0.0163 (0.0015)
σ	0.2143 (0.0243)	0.1954 (0.0198)	0.2142 (0.0208)	0.2284 (0.0251)	0.1906 (0.0222)	0.1948 (0.0208)	0.1941 (0.0197)	0.2147 (0.0208)	0.2105 (0.0237)	0.1722 (0.0202)
α	0.7565 (0.0603)	0.7825 (0.0558)	0.8440 (0.0620)	0.7929 (0.0627)	0.6039 (0.0390)	0.7027 (0.0470)	0.7766 (0.0541)	0.8403 (0.0602)	0.7401 (0.0569)	0.5674 (0.0335)

Table 4: Linear regression model results for Boston dataset. Standard deviations in parenthesis.

distribution all the variables appear in the table with a sign in line with previous studies on the same dataset. Nevertheless, Lasso prior should be preferred at least for two reasons. First it uniformly provides smaller posterior standard errors, the estimated coefficients appear to be more reliable at extreme quantile levels, i.e. $\tau = 0.1$ or $\tau = 0.9$ for which the estimated parameters obtained under Gaussian prior choice become very unstable for some variables.

6.2. Munich rental guide

In section 5.3 we provide empirical evidence that the SEP distribution produces more reliable estimates of the conditional quantile in presence of heteroskedasticity and heavy tails. To provide a real data example we analyze the very well known 2003 Munich rental dataset, notoriously characterized by the presence of heterogeneous variability. Furthermore, several analyses of this dataset (see for example Kneib et al, 2011 and Mayr et al, 2012) showed the presence of a

spatial effects modeled by considering a parameter for each of the 380 districts of Munich. For this reason the parameter space handled is quite wide highlighting the need of considering a variable selection approach. Here therefore, we assume a Lasso prior distribution on the unknown parameters in line with the one proposed in 28 and we compare its performances with a Gaussian prior assumption. The response variable is the rent in Euro per square meters for a flat in Munich. Two sets of covariates describe linear and non linear relations between the rent and its determinants. The linear predictors are a set of 13 dummies for the goodness of location, the goodness of rooms and the number of rooms in the flat. The floor size and the year of construction have instead a non linear impact on the response variable. Finally, the spatial location of the flat allows to implement a geoaddivitive model of the kind introduced by Kammann and Wand (2003). To this aim we use a Bayesian semi-parametric quantile regression model with a spatial effect similar to the one considered in Rue and Held (2005) and Yue and Rue (2011) among the others. A complete description of the dataset can be found in Rue and Held (2005).

We estimate the τ -th conditional quantile for the rent r_t , i.e., $Q_\tau(r_t \mid \mathbf{x}_t, \mathbf{z}_t)$ from the following model:

$$\begin{aligned} r_t &= q_{t,\tau} + \epsilon_t \\ q_{t,\tau} &= \mathbf{x}_t' \boldsymbol{\beta}_\tau + f_{s,\tau}(z_{s,t}) + f_{y,\tau}(z_{y,t}) + f_{l,\tau}(z_{l,t}), \end{aligned} \quad (50)$$

where $t = 1, 2, \dots, 2035$; ϵ_t is the error term with zero τ -th quantile and constant variance, \mathbf{x}_t is the whole set of dummies treated as linear parametric predictors, $\mathbf{z}_t = (z_{s,t}, z_{y,t}, z_{l,t})$ are the predictor variable for “size”, “year” and “spatial” effect while $f_{s,\tau}$, $f_{y,\tau}$ and $f_{l,\tau}$ are their non linear functions. The estimation procedure of three quantile confidence levels $\tau = (0.25, 0.5, 0.75)$ has been performed using the Adaptive MCMC procedure for GAM models described in subsection 4.2.

Figures 3 and 4 show the estimated non linear effect for the year of construction and the floor size using Gaussian and Lasso prior respectively. The points below each sub-figure represent the available observations for each value of the covariates while dotted lines represent the 95% posterior credible intervals. We can observe that both priors provide similar estimated splines for the effect of the floor size on the house prices. It can be seen that small flat (less than $40m^2$) has a very high rent for square meters while for big flat the rent remains almost unchanged. Concerning the year of construction, the estimated splines are apparently quite different under the two prior specifications but they actually contain similar information. In fact, looking at the level of the variable and the confidence intervals the effect of this covariate can be approximatively considered equal to zero until the 1990, when a clear positive and increasing effect is shown under both prior specification. Figure 5 displays the estimated spatial effects of 380 subquarters in Munich. Values are normalized to be in the range $(0, 1)$. As expected, for both Gaussian and Lasso prior, rents are high in the centre of Munich and some well-known districts, while it becomes lower on the margins. Finally, estimated posterior means and standard deviations for the

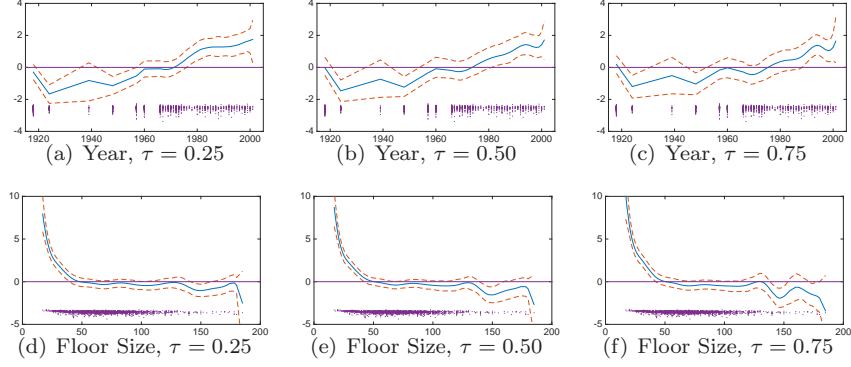


Figure 3: Estimated nonparametric effect using Gaussian prior with 95% credible bands for Munich data.

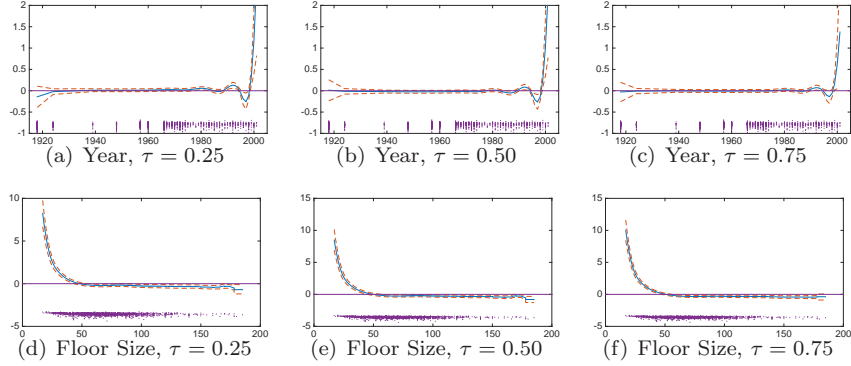


Figure 4: Estimated nonparametric effect using Lasso prior with 95% credible bands for Munich data.

linear parametric effects are shown in Table 5. The signs of the variables are in line with previous works but new interesting results are suggested under Lasso prior in the estimation of the effect of "No hot water", "No central heating" and "6 Rooms". In particular, Lasso prior discriminates further the effect of these variables for each quantile. Indeed, we can see that the absence of hot water and the presence of 6 Rooms have a statistical significant effect only on expensive house, i.e. for $\tau = 0.75$ while the opposite occurs considering the absence of central heating. We think that these results highlight more consistently the variety of the consumption choices due to different budget constraints. It is worth noting that Lasso prior correctly shrinks the effect of "Special bathroom interior" that is not very significant when estimated using Gaussian prior.

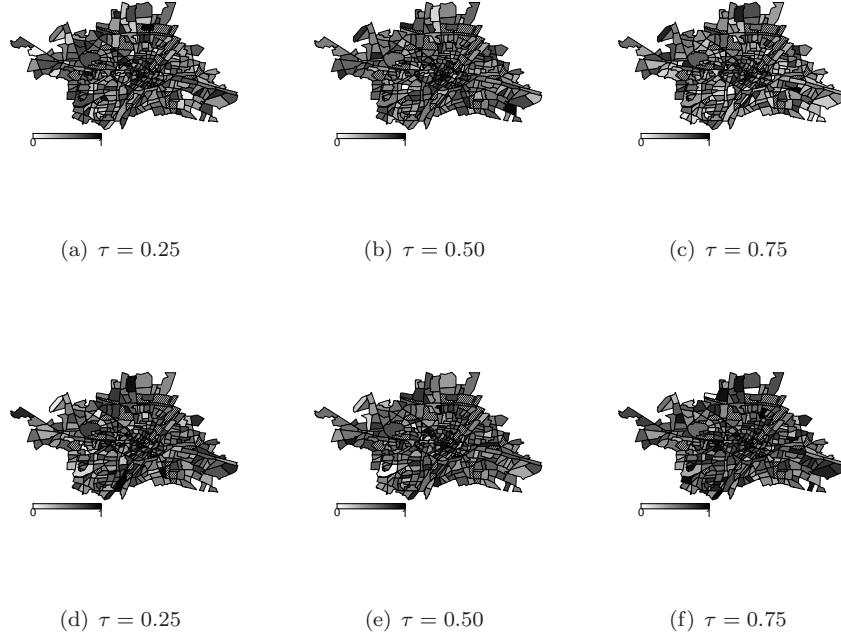


Figure 5: Estimated spatial effects using Gaussian (first row) and Lasso (second row) prior for the 380 subquarters of Munich

6.3. Barro growth data

As final application, we analyze the dataset related to the international economic growth model firstly considered by Barro and Sala i-Martin (1995) and extended to the quantile regression framework by Koenker and Machado (1999). Since standard OLS model do not provide a clear result about the convergence hypothesis of neoclassical growth models, several papers have analyzed growth equations using quantile regression technique with prominent results. In their paper Barreto and Hughes (2004) show that the determinants of the economic growth for countries in the left or right tails of the distribution are very different from those in the mean. Mello and Perrelli (2003) use quantile regression to find evidence in favor of the convergence hypothesis for countries in the upper quantile of the conditional distribution of the response variable using the Barro growth model (Barro, 1991). Finally, Laurini (2007) uses spline functions in testing the convergence hypothesis with a dataset of Brazilian municipalities. To the best of our knowledge, this is the first attempt to propose a Bayesian quantile Lasso GAM model in order to study the impact of both linear and non linear effect of the covariates on the cross country GDP growth using the Barro and Sala i-Martin (1995) model. The dataset contains 161 world nations

Variable	Gaussian Prior			Lasso Prior		
	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Good location	0.6466 (0.0925)	0.7454 (0.0880)	0.7606 (0.0857)	0.6304 (0.1230)	0.7042 (0.1124)	0.5922 (0.1039)
Excellent location	1.4213 (0.2770)	1.6999 (0.2879)	1.9381 (0.2829)	1.4136 (0.2376)	1.6305 (0.2454)	1.8450 (0.2527)
No hot water	-1.3361 (0.2410)	-1.8499 (0.2664)	-2.2199 (0.2738)	-0.0353 (0.0475)	-0.0335 (0.0454)	-2.7652 (0.3731)
No central heating	-1.5449 (0.1759)	-1.4206 (0.1957)	-1.0610 (0.1867)	-1.9830 (0.1872)	-2.0557 (0.2076)	-0.1316 (0.1193)
No tiles in bathroom	-0.4260 (0.1079)	-0.5792 (0.1091)	-0.5942 (0.1140)	-0.1597 (0.1143)	-0.3277 (0.1453)	-0.2576 (0.1575)
Special bathroom interior	0.3926 (0.1462)	0.3803 (0.1580)	0.4897 (0.1489)	0.0598 (0.0606)	0.0824 (0.0779)	0.0598 (0.0581)
Special kitchen interior	0.9145 (0.1787)	1.1405 (0.1564)	1.2480 (0.1740)	1.0824 (0.3355)	1.3077 (0.2239)	1.3153 (0.1989)
1 Room	7.1564 (0.1754)	8.3633 (0.1700)	9.3637 (0.1751)	6.8372 (0.1926)	8.6886 (0.1729)	10.0425 (0.1652)
2 Rooms	6.9968 (0.1002)	8.4530 (0.1024)	10.0062 (0.0922)	6.5432 (0.1154)	8.5664 (0.1126)	10.3823 (0.1084)
3 Rooms	6.7542 (0.0998)	8.1964 (0.0927)	9.7554 (0.0947)	6.2149 (0.1159)	8.1500 (0.1101)	10.0117 (0.1076)
4 Rooms	6.2745 (0.1459)	7.7603 (0.1404)	9.2060 (0.1490)	5.7041 (0.1557)	7.5529 (0.1644)	9.3650 (0.1601)
5 Rooms	6.0948 (0.2659)	7.6821 (0.3047)	9.6398 (0.2956)	5.0121 (0.4154)	7.0744 (0.3538)	9.2734 (0.3514)
6 Rooms	6.3496 (0.4450)	7.6293 (0.4479)	9.1707 (0.4661)	0.4008 (0.0658)	0.4462 (0.0752)	8.4068 (0.5960)

Table 5: Posterior Means and standard errors (in parenthesis) of the linear regressors for three different quantile levels.

observed for 13 covariates covering the two periods 1965-75 and 1975-85. With a quantile GAM model we are able to combine the theory of non linear return to education with that of economic convergence using spline functions to model the variables "Male secondary school" (MSS), "Female Secondary school" (FSS), "Male Higher Education" (MHE) and "Female Higher Education" (FHE) while we adopt a linear representation for the remaining variables.

The parameter estimates of the linear covariates (Table 6) are in line with previous studies based on quantile regression methods. In particular, it is worth noting that the coefficients related to the initial per capita GDP is always negative, confirming the neoclassical theory about conditional convergence. Figures 6 displays the estimated spline functions along with their credible sets, for three quantile levels $\tau = (0.25, 0.5, 0.75)$. A noticeable non linear path is showed for almost all the selected covariates. For a given variable, the sign of each estimated spline varies among different quantile levels suggesting that the importance of

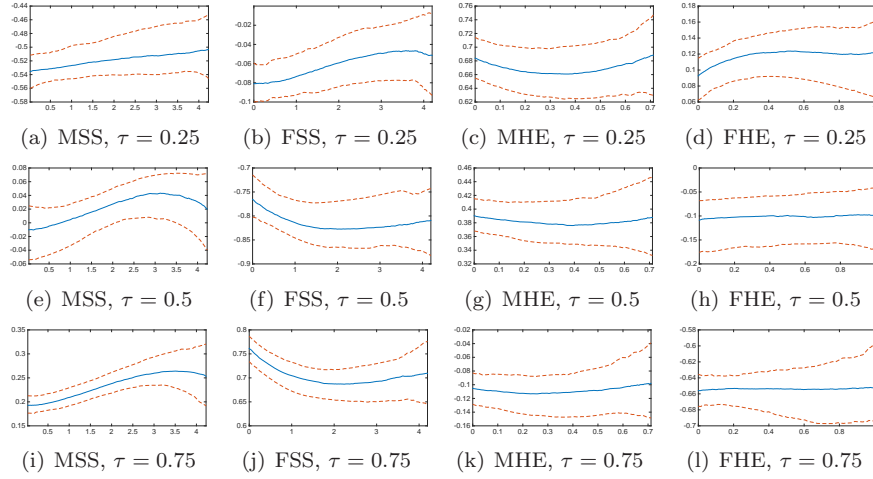


Figure 6: Barro dataset. Penalized splines with 95% HPD credible sets for the variables: “Male secondary school” (MSS, first column), “Female secondary school ” (FSS, second column), “Male Higher education ” (MHE, third column) and Female Higher education ” (FHE, fourth column) for five different quantile levels.

different types of education is not the same for countries in the lower and upper tails of the growth conditional distribution. This result is of particular interest since it allows to isolate the positive and negative contributions of each type of education on the rate of economic growth. There are two opposite paths characterizing the effect of secondary schooling and higher education on growth: the first one is increasing in the quantile level τ , the second is decreasing. In particular our estimates suggest that relatively low education levels help countries in the upper tail of the growth distribution while higher education levels boost the rate of growth for countries in the lower tail. Those results can be interpreted in view of the fact that, high and low GDP growth levels are linked with emerging and developed nations respectively. The basic schooling is a key factor for emerging nations which base their economies on high labor intensity activities. For these countries, the costs resulting from higher levels of schooling outweigh their returns while the opposite is true for the advanced countries. The latter, indeed, have the possibility to take advantage of skilled labor forces to exploit the higher returns derived from the available technology.

7. Conclusion

In this paper we show how the SEP distribution provides a flexible tool to model the conditional quantile of a response variable as a function of exogenous covariates in a Bayesian quantile regression contest. In particular extreme observations are properly accounted by the shape parameter governing the tails decay of the distribution efficiently handling data with outliers or with fat tail-decay. Moreover we extend the linear quantile regression framework to the

Variable	Quantile levels				
	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Initial Per Capita GDP	-0.0266 (0.0043)	-0.0270 (0.0048)	-0.0307 (0.0055)	-0.0293 (0.0045)	-0.0322 (0.0047)
Life Expectancy	0.0324 (0.0091)	0.0127 (0.0104)	0.1947 (0.0110)	0.2021 (0.0097)	0.1866 (0.0092)
Human Capital	-0.0024 (0.0011)	-0.0037 (0.0017)	-0.0010 (0.0020)	-0.0023 (0.0019)	-0.0018 (0.0017)
Education/GDP	-0.2707 (0.1246)	-0.1127 (0.1579)	-0.2956 (0.1690)	-0.2156 (0.1721)	-0.0993 (0.1732)
Investment/GDP	0.0999 (0.0268)	0.0930 (0.0291)	0.0359 (0.0337)	0.0343 (0.0277)	0.0482 (0.0264)
Public Consumption/GDP	-0.1628 (0.0387)	-0.1723 (0.0443)	0.0052 (0.0463)	0.0251 (0.0384)	-0.0206 (0.0280)
Black Market Premium	-0.0227 (0.0063)	-0.0267 (0.0074)	-0.0360 (0.0077)	-0.0319 (0.0064)	-0.0316 (0.0072)
Political Instability	-0.0264 (0.0083)	-0.0302 (0.0088)	-0.0153 (0.0103)	-0.0053 (0.0098)	-0.0042 (0.0073)
Growth Rate Terms Trade	0.1220 (0.0366)	0.1250 (0.0504)	0.2274 (0.0652)	0.2478 (0.0639)	0.2744 (0.0569)

Table 6: Posterior Means and standard errors (in parenthesis) of the linear regressors for three different quantile levels.

GAM one when quantile functions are approximated with splines. In both cases we provide new adaptive Metropolis within Gibbs algorithm in order to implement the statistical inference. Since it is common when building models, that a big number of parameters should be estimated in particular when spline tools are used, in this paper we accommodate the problem of variable selection and shrinking parameters by using the Bayesian version of Lasso penalization methods. In particular we suggest the use of generalized independent Laplace priors on the regressor parameters in the linear case allowing to shrink each parameter separately and a multivariate Laplace distribution on the spline coefficients generalizing the Lang and Brezger (2004) second order random walk prior. Finally we show the power of the models considered through simulation and real data set applications where it is evident the flexibility of the quantile methodology proposed in terms of robustness and sparsity.

Acknowledgments

This research is supported by the Italian Ministry of Research PRIN 2013–2015, “Multivariate Statistical Methods for Risk Assessment” (MISURA), and by the “Carlo Giannini Research Fellowship”, the “Centro Interuniversitario di Econometria” (CIdE) and “UniCredit Foundation”.

Appendix A

Lemma 7.1. *Let $Y \sim \mathcal{SEP}(\mu, \sigma, \alpha, \tau)$, then the τ -level quantile of Y coincides with its natural location parameter, i.e. $Q_\tau(Y) = \mu$.*

Proof. In order to show that $P(Y \leq \mu) = \tau$ we compute the cdf of a SEP in $y = \mu$

$$\begin{aligned} P(Y \leq \mu) &= \int_{-\infty}^{\mu} \frac{1}{2\sigma} \frac{\alpha^{-\frac{1}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})} \left[\exp \left\{ -\frac{1}{\alpha} \left(\frac{\mu - y}{2\tau\sigma} \right)^\alpha \right\} \mathbb{1}_{(y \leq \mu)} \right. \\ &\quad \left. + \exp \left\{ -\frac{1}{\alpha} \left(\frac{y - \mu}{2(1-\tau)\sigma} \right)^\alpha \right\} \mathbb{1}_{(y > \mu)} \right] dy \\ &= \int_{-\infty}^{\mu} \frac{1}{2\sigma} \frac{\alpha^{-\frac{1}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})} \exp \left\{ -\frac{1}{\alpha} \left(\frac{\mu - y}{2\tau\sigma} \right)^\alpha \right\} dy \end{aligned} \quad (51)$$

Without loss of generality, let us consider the case when $\mu = 0$ and $\sigma = 1$. The integral reduces to

$$\frac{\alpha^{-\frac{1}{\alpha}}}{2\Gamma(1 + \frac{1}{\alpha})} \int_{-\infty}^0 \exp \left\{ -\frac{1}{\alpha} \left(\frac{-y}{2\tau} \right)^\alpha \right\} dy. \quad (52)$$

By substitute $(-y)^\alpha = x$ we have

$$\frac{\alpha^{-\frac{1}{\alpha}}}{2\Gamma(1 + \frac{1}{\alpha})} \int_0^{+\infty} \exp \left\{ -\frac{x}{\alpha(2\tau)^\alpha} \right\} \frac{1}{\alpha} (x)^{\frac{1}{\alpha}-1} dx \quad (53)$$

Rearranging equation (53) and recognizing the kernel of a Gamma pdf with shape $1/\alpha$ and scale $\alpha(2\tau)^\alpha$ the integral becomes

$$\frac{\alpha^{-\frac{1}{\alpha}}}{2\Gamma(1 + \frac{1}{\alpha})} \frac{1}{\alpha} \Gamma\left(\frac{1}{\alpha}\right) (\alpha(2\tau)^\alpha)^{\frac{1}{\alpha}}.$$

By using the property $\Gamma(x+1) = x\Gamma(x)$ all the terms simplify except for τ , concluding the proof. \square

References

- Alhamzawi, R., Yu, K., and Benoit, D. F. (2012). Bayesian adaptive lasso quantile regression. *Statistical Modelling*, 12(3):279–297.
- Andrieu, C. and Moulines, r. (2006). On the ergodicity properties of some adaptive mcmc algorithms. *The Annals of Applied Probability*, 16(3):1462–1505.
- Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive mcmc. *Statistics and Computing*, 18(4):343–373.
- Barreto, R. A. and Hughes, A. W. (2004). Under performers and over achievers: A quantile regression analysis of growth. *Economic Record*, 80(248):17–35.
- Barro, R. and i Martin, X. S. (1995). *Economic Growth*. McGraw-Hill, New York.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106(2):407–443.
- Bernardi, M., Gayraud, G., and Petrella, L. (2015). Bayesian tail risk interdependence using quantile regression. *Bayesian Analysis*, 10(3):pp. 553–603.
- Box, G. and Tiao, G. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley series in behavioral science: quantitative methods. Addison-Wesley Pub. Co.
- Brezger, A. and Lang, S. (2006). Generalized structured additive regression based on bayesian p-splines. *Computational Statistics and Data Analysis*, 50(4):967 – 991.
- Brezger, A. and Steiner, W. J. (2008). Monotonic regression based on bayesian p-splines: An application to estimating price response functions from store-level scanner data. *Journal of Business and Economic Statistics*, 26(1):pp. 90–104.
- Chen, C. and Yu, K. (2009). Automatic bayesian quantile regression curve fitting. *Statistics and Computing*, 19(3):271–281.
- Choy, S. and Walker, S. G. (2003). The extended exponential power distribution and bayesian robustness. *Statistics & Probability Letters*, 65(3):227 – 232.
- Choy, S. T. B. and Smith, A. F. M. (1997). On robust analysis of a normal location parameter. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(2):pp. 463–474.
- Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998). Automatic bayesian curve fitting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):333–350.

- Dunson, D. B. and Taylor, J. A. (2005). Approximate bayesian inference for quantiles. *Journal of Nonparametric Statistics*, 17(3):385–400.
- Eilers, P. H. C., Rijnmond, D. M., and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121.
- Fernandez, C. and Steel, M. F. J. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):pp. 359–371.
- Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4):835–845.
- Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3):pp. 297–310.
- Jr., D. H. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1):81 – 102.
- Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1):1–18.
- Kobayashi, G. (2016). Skew exponential power stochastic volatility model for analysis of skewness, non-normal tails, quantiles and expectiles. *Computational Statistics*, 31(1):49–88.
- Koenker, B. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.
- Koenker, B. and Basset, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.
- Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310.
- Kottas, A. and Gelfand, A. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96:1458–1468.
- Kottas, A. and Krnjajic, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, 36:297–319.
- Kotz, S., Kozubowski, T. J., and Podgórski, K. (2001). Asymmetric multivariate laplace distribution. In *The Laplace Distribution and Generalizations*, pages 239–272. Springer.
- Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81:1565–1578.

- Lang, S. and Brezger, A. (2004). Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13(1):pp. 183–212.
- Laurini, M. P. (2007). A note on the use of quantile regression in beta convergence analysis. *Economics Bulletin*, 3(52):1–8.
- Li, Q., Xi, R., and Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Anal.*, 5(3):533–556.
- Lum, K. and Gelfand, A. (2012). Spatial quantile multiple regression using the asymmetric laplace process. *Bayesian Analysis*, 7:235–258.
- Mayr, A., Fenske, N., Hofner, B., Kneib, T., and Schmid, M. (2012). Generalized additive models for location, scale and shape for high dimensional data: a flexible approach based on boosting. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):403–427.
- Meier, L., Van De Geer, S., and Bhlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71.
- Mello, M. and Perrelli, R. (2003). Growth equations: a quantile regression exploration. *The Quarterly Review of Economics and Finance*, 43(4):643 – 667. Capital Accumulation and Allocation in Economic Growth.
- Naranjo, L., Pérez, C. J., and Martín, J. (2015). Bayesian analysis of some models that use the asymmetric exponential power distribution. *Statistics and Computing*, 25(3):497–514.
- Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847.
- O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Science*, 1:505–527.
- OSullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing*, 9(2):363–379.
- Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103:681–686.
- Roberts, G. O. and Rosenthal, J. S. (2007). Coupling and ergodicity of adaptive markov chain monte carlo algorithms. *Journal of Applied Probability*, 44(2):pp. 458–475.
- Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.
- Salazar, E., Ferreira, M., and Migon, H. (2012). Objective bayesian analysis for exponential power regression models. *Sankhya B*, 74(1):107–125.

- Sriram, K., Ramamoorthi, R., and Ghosh, P. (2013). Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Analysis*, 8:479–504.
- Thomas Kneib, Susanne Konrath, L. F. (2011). High dimensional structured additive regression models: Bayesian regularization, smoothing and predictive performance. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 60(1):51–70.
- Thompson, P., Cai, Y., Moyeed, R., Reeve, D., and Stander, J. (2010). Bayesian nonparametric quantile regression using splines. *Computational Statistics and Data Analysis*, 54(4):1138 – 1150.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288.
- Tokdar, S. and Kadane, J. (2012). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Economics*, 6:382–406.
- Wichitaksorn, N., Choy, S. T. B., and Gerlach, R. (2014). A generalized class of skew distributions and associated robust quantile regression models. *Canadian Journal of Statistics*, 42(4):579–596.
- Yu, K. and Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54:437–447.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.
- Yue, Y. R. and Rue, H. (2011). Bayesian inference for additive mixed quantile regression models. *Comput. Stat. Data Anal.*, 55(1):84–96.
- Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, 148(1):86 – 99.